

08/31/9, 831A

Notice of Availability

Applicant Aid for Biotechnology Computer Readable Form (CRF) Sequence Listings Submissions

The Patent and Trademark Office (PTO) has developed a computer program, called Checker, that will aid applicants in identifying and correcting errors prior to making submissions for compliance with the Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence Disclosures (sequence rules: 37 CFR 1.821 through 1.825). (Final rules were published in the *Federal Register* (55 FR 18230) on May 1, 1990, and in the *PTO Official Gazette* (1114 Off.Gaz.PatOffice 29) on May 15, 1990.)

Checker is a DOS-based software program that is intended to assist users in determining whether errors may be present in the sequence listings, and is not intended to guarantee that the submission is error-free.

The most current version of the software will be available via computer downloading (details below). Copies on diskette are also available. Updated software versions will not be automatically mailed out; any updates will be announced in the *PTO Official Gazette*.

The software can be accessed/requested in the following locations:

- 1) Dial-up access to the Patent and Trademark Office Bulletin Board System.
Phone number: 703-305-8950
Cost: Free-of-charge
- 2) Dial-up access through the Internet. FTP site: ftp.uspto.gov
Login as "anonymous". Software is in directory /pub/checker
Cost: Free-of-charge
- 3) For diskette copies, telephone requests to 703-306-2600.
Cost: \$25.00

For Further Information Contact: Meredith Beckhardt at 703-308-4212.

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:36

INPUT SET: S9275.raw

This Raw Listing contains the General
Information Section and those Sequences
containing ERRORS.

1
2
3 (1) General Information:
4
5 (i) APPLICANT: Hewick, Rodney M.
6 Wang, Jack H.
7 Wozney, John M.
8 Celeste, Anthony J.
9
10 (ii) TITLE OF INVENTION: BONE AND CARTILAGE INDUCTIVE PROTEINS
11
12 (iii) NUMBER OF SEQUENCES: 44
13
14 (iv) CORRESPONDENCE ADDRESS:
15 (A) ADDRESSEE: Legal Affairs - Genetics Institute, Inc.
16 (B) STREET: 87 CambridgePark Drive
17 (C) CITY: Cambridge
18 (D) STATE: MA
19 (E) COUNTRY: USA
20 (F) ZIP: 02140
21
22 (v) COMPUTER READABLE FORM:
23 (A) MEDIUM TYPE: Floppy disk
24 (B) COMPUTER: IBM PC compatible
25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
26 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
27
28 (vi) CURRENT APPLICATION DATA:
29 (A) APPLICATION NUMBER: US 08/319,831
30 (B) FILING DATE: 06-OCT-1994
31 (C) CLASSIFICATION:
32
33 (viii) ATTORNEY/AGENT INFORMATION:
34 (A) NAME: Kapinos, Ellen J.
35 (B) REGISTRATION NUMBER: 32,245
36 (C) REFERENCE/DOCKET NUMBER: GI 5182A-DIV
37
38 (ix) TELECOMMUNICATION INFORMATION:
39 (A) TELEPHONE: (617) 498-8622
40 (B) TELEFAX: (617) 876-5851
41

Does Not Comply
Corrected Diskette Needed

ERRORED SEQUENCES FOLLOW:

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:38

INPUT SET: S9275.raw

144 (2) INFORMATION FOR SEQ ID NO:5:
145
146 (i) SEQUENCE CHARACTERISTICS:
--> 147 (A) LENGTH: 80 base pairs
148 (B) TYPE: nucleic acid
149 (C) STRANDEDNESS: double
150 (D) TOPOLOGY: linear
151
152 (ii) MOLECULE TYPE: DNA (genomic)
153
154 (iii) HYPOTHETICAL: NO
155
156 (iv) ANTI-SENSE: NO
157
158 (vi) ORIGINAL SOURCE:
159 (A) ORGANISM: Bos taurus
160
161 (vii) IMMEDIATE SOURCE:
162 (B) CLONE: acc30
163
164 (viii) POSITION IN GENOME:
165 (C) UNITS: bp
166
167 (ix) FEATURE:
168 (A) NAME/KEY: CDS
169 (B) LOCATION: 25..57
170
171 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
172
173 GGATCCGCGT GCTGTGCTCC GACC AAG CTG AGC GCC ACC TCC GTG CTC TAC
174 51
175 Lys Leu Ser Ala Thr Ser Val Leu Tyr
176 1 5
177
--> 178 TAC GAC AGCAGCAACA ATGTAATTC T AGA 80
179 Tyr Asp
180 10
181

move up to this line

196 (2) INFORMATION FOR SEQ ID NO:7:
197
198 (i) SEQUENCE CHARACTERISTICS:
--> 199 (A) LENGTH: 199 base pairs
200 (B) TYPE: nucleic acid
201 (C) STRANDEDNESS: double
202 (D) TOPOLOGY: linear
203
204 (ii) MOLECULE TYPE: DNA (genomic)
205
206 (iii) HYPOTHETICAL: NO
207
208 (vi) ORIGINAL SOURCE:
209 (A) ORGANISM: Bos taurus

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:40

INPUT SET: S9275.raw

210
211 (vii) IMMEDIATE SOURCE:
212 (A) LIBRARY: Bovine genomic
213 (B) CLONE: Lambda 9800-10
214
215 (viii) POSITION IN GENOME:
216 (C) UNITS: bp
217
218 (ix) FEATURE:
219 (A) NAME/KEY: exon
220 (B) LOCATION: 30..199
221
222 (ix) FEATURE:
223 (A) NAME/KEY: intron
224 (B) LOCATION: 1..29
225
226 (ix) FEATURE:
227 (A) NAME/KEY: CDS
228 (B) LOCATION: 30..179
229
230 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

--> 231
232 TGCCCCGCTGC CCCCTCCCGC CCCC GCCAG GTG CAC CTG CTG AAG CCG CAC
233 GCG 53
234
235
236
237 GTC CCC AAG GCG TGC TGC GCG CCC ACC AAG CTG AGC GCC ACT TCC GTG
238 101
239 Val Pro Lys Ala Cys Cys Ala Pro Thr Lys Leu Ser Ala Thr Ser Val
240 10 15 20 25
241
242 CTC TAC TAC GAC AGC AGC AAC AAC GTC ATC CTG CGC AAG CAC CGC AAC
243 149
244 Leu Tyr Tyr Asp Ser Ser Asn Asn Val Ile Leu Arg Lys His Arg Asn
245 30 35 40
246
247 ATG GTG GTC CGC GCC TGC GGC TGC CAC TGA GGCCCCAACT CCACCGGCAG
248 199
249 Met Val Val Arg Ala Cys Gly Cys His
250 45 50
251

once this total is
moved up one line,
amino acids should
be aligned under codons
Val His Leu Leu Lys
1

252 (2) INFORMATION FOR SEQ ID NO:8:

253

254 (i) SEQUENCE CHARACTERISTICS:

--> 255 (A) LENGTH: 50 amino acids

256 (B) TYPE: amino acid

257 (D) TOPOLOGY: linear

258

259 (ii) MOLECULE TYPE: protein

260

261 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

49 are shown - is one
missing?

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:43

INPUT SET: S9275.raw

262
263 Val His Leu Leu Lys Pro His Ala Val Pro Lys Ala Cys Cys Ala Pro
264 1 5 10 15
265
266 Thr Lys Leu Ser Ala Thr Ser Val Leu Tyr Tyr Asp Ser Ser Asn Asn
267 20 25 30
268
269 Val Ile Leu Arg Lys His Arg Asn Met Val Val Arg Ala Cys Gly Cys
270 35 40 45
271
272 His
273
274
275
276

277 (2) INFORMATION FOR SEQ ID NO:9:
278
279 (i) SEQUENCE CHARACTERISTICS:
--> 280 (A) LENGTH: 172 base pairs
281 (B) TYPE: nucleic acid
282 (C) STRANDEDNESS: double
283 (D) TOPOLOGY: linear
284
285 (ii) MOLECULE TYPE: DNA (genomic)
286
287 (iii) HYPOTHETICAL: NO
288
289 (vi) ORIGINAL SOURCE:
290 (A) ORGANISM: Bos taurus
291
292 (vii) IMMEDIATE SOURCE:
293 (A) LIBRARY: Bovine genomic
294 (B) CLONE: Lambda 9800-10
295
296 (viii) POSITION IN GENOME:
297 (C) UNITS: bp
298
299 (ix) FEATURE:
300 (A) NAME/KEY: exon
301 (B) LOCATION: 51..161
302
303 (ix) FEATURE:
304 (A) NAME/KEY: intron
305 (B) LOCATION: 1..50
306
307 (ix) FEATURE:
308 (A) NAME/KEY: intron
309 (B) LOCATION: 162..172
310
311 (ix) FEATURE:
312 (A) NAME/KEY: CDS
313 (B) LOCATION: 51..161

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:45

INPUT SET: S9275.raw

314
315 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
316
317 GGGGTGGGAG GGCACGTGGA TGGGACTCAC CTTCTCCCAC TACCCCCCAG GAC
--> 318 TGG 56
319 Asp Trp
320 1
321 GTC ATC GCC CCC CAA GGC TAC TCA GCC TAT TAC TGT GAA GGG GAG TGC
322 104
323 Val Ile Ala Pro Gln Gly Tyr Ser Ala Tyr Tyr Cys Glu Gly Glu Cys
324 5 10 15
325
326 TCC TTC CCG CTG GAC TCC TGC ATG AAC GCC ACC AAC CAC GCC ATC CTG
327 152
328 Ser Phe Pro Leu Asp Ser Cys Met Asn Ala Thr Asn His Ala Ile Leu
329 20 25 30
330
--> 331 CAG TCC CTG GTCAGTACCT C 172
332 Gln Ser Leu
333 5
334

355 (2) INFORMATION FOR SEQ ID NO:11:
356
357 (i) SEQUENCE CHARACTERISTICS:
--> 358 (A) LENGTH: 119 base pairs
359 (B) TYPE: nucleic acid
360 (C) STRANDEDNESS: double
361 (D) TOPOLOGY: linear
362
363 (ii) MOLECULE TYPE: DNA (genomic)
364
365 (iii) HYPOTHETICAL: NO
366
367 (vi) ORIGINAL SOURCE:
368 (A) ORGANISM: Bos taurus
369
370 (vii) IMMEDIATE SOURCE:
371 (A) LIBRARY: Bovine genous
372 (B) CLONE: Lambda 9800-10
373
374 (viii) POSITION IN GENOME:
375 (C) UNITS: bp
376
377 (ix) FEATURE:
378 (A) NAME/KEY: exon
379 (B) LOCATION: 20..99
380
381 (ix) FEATURE:
382 (A) NAME/KEY: intron
383 (B) LOCATION: 1..19
384
385 (ix) FEATURE:

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:47

INPUT SET: S9275.raw

386 (A) NAME/KEY: intron
387 (B) LOCATION: 100..119
388
389 (ix) FEATURE:
390 (A) NAME/KEY: CDS
391 (B) LOCATION: 22..99
392
393
394 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
395
396 CCCTTGCGTG TCCCCGCAGA C GAC GTC CAC GGC TCC CAC GGC CGG CAG
--> 397 GTG 51
398 Asp Val His Gly Ser His Gly Arg Gln Val
399 1 5 10
400
401 TGC CGT CGG CAC GAG CTG TAC GTG AGC TTC CAG GAC CTG GGC TGG CTG
402 99
403 Cys Arg Arg His Glu Leu Tyr Val Ser Phe Gln Asp Leu Gly Trp Leu
404 15 20 25
405
--> 406 GTGAGTTCCG ACTCTCCTTT
407

425 (2) INFORMATION FOR SEQ ID NO:13:
426
427 (i) SEQUENCE CHARACTERISTICS:
--> 428 (A) LENGTH: 1003 base pairs
429 (B) TYPE: nucleic acid
430 (C) STRANDEDNESS: double
431 (D) TOPOLOGY: circular
432
433 (ii) MOLECULE TYPE: cDNA to mRNA
434
435 (iii) HYPOTHETICAL: NO
436
437 (vi) ORIGINAL SOURCE:
438 (A) ORGANISM: Homo sapiens
439 (F) TISSUE TYPE: Human Heart
440
441 (vii) IMMEDIATE SOURCE:
442 (A) LIBRARY: Human heart cDNA library stratagene catalog
443 (B) CLONE: hh38
444
445 (viii) POSITION IN GENOME:
446 (C) UNITS: bp
447
448 (ix) FEATURE:
449 (A) NAME/KEY: CDS
450 (B) LOCATION: 8..850
451
452 (ix) FEATURE:
453 (A) NAME/KEY: mat_peptide
454 (B) LOCATION: 427..843

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:50

INPUT SET: S9275.raw

455

456

(ix) FEATURE:

457

(A) NAME/KEY: mRNA

458

(B) LOCATION: 1..997

459

460

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

461

462

GAATTCC GAG CCC CAT TGG AAG GAG TTC CGC TTT GAC CTG ACC CAG ATC

463

49

464

Glu Pro His Trp Lys Glu Phe Arg Phe Asp Leu Thr Gln Ile

465

-139

-135

-130

466

467

CCG GCT GGG GAG GCG GTC ACA GCT GCG GAG TTC CGG ATT TAC AAG GTG

468

97

469

Pro Ala Gly Glu Ala Val Thr Ala Ala Glu Phe Arg Ile Tyr Lys Val

470

-125

-120

-115

-110

471

472

CCC AGC ATC CAC CTG CTC AAC AGG ACC CTC CAC GTC AGC ATG TTC CAG

473

145

474

Pro Ser Ile His Leu Leu Asn Arg Thr Leu His Val Ser Met Phe Gln

475

-105

-100

-95

476

477

GTG GTC CAG GAG CAG TCC AAC AGG GAG TCT GAC TTG TTC TTT TTG GAT

478

193

479

Val Val Gln Glu Gln Ser Asn Arg Glu Ser Asp Leu Phe Phe Leu Asp

480

-90

-85

-80

481

482

CTT CAG ACG CTC CGA GCT GGA GAC GAG GGC TGG CTG GTG CTG GAT GTC

483

241

484

Leu Gln Thr Leu Arg Ala Gly Asp Glu Gly Trp Leu Val Leu Asp Val

485

-75

-70

-65

486

487

ACA GCA GCC AGT GAC TGC TGG TTG CTG AAG CGT CAC AAG GAC CTG GGA

488

289

489

Thr Ala Ala Ser Asp Cys Trp Leu Leu Lys Arg His Lys Asp Leu Gly

490

-60

-55

-50

491

492

CTC CGC CTC TAT GTG GAG ACT GAG GAT GGG CAC AGC GTG GAT CCT GGC

493

337

494

Leu Arg Leu Tyr Val Glu Thr Glu Asp Gly His Ser Val Asp Pro Gly

495

-45

-40

-35

-30

496

497

CTG GCC GGC CTG CTG GGT CAA CGG GCC CCA CGC TCC CAA CAG CCT TTC

498

385

499

Leu Ala Gly Leu Leu Gly Gln Arg Ala Pro Arg Ser Gln Gln Pro Phe

500

-25

-20

-15

501

502

GTG GTC ACT TTC TTC AGG GCC AGT CCG AGT CCC ATC CGC ACC CCT CGG

503

433

504

Val Val Thr Phe Phe Arg Ala Ser Pro Ser Pro Ile Arg Thr Pro Arg

505

-10

-5

1

506

507

*edit
throughout*

RAW SEQUENCE LISTING

PATENT APPLICATION US/08/319,831A

DATE: 05/09/96
TIME: 16:30:52

INPUT SET: S9275.raw

```

508 GCA GTG AGG CCA CTG AGG AGG AGG CAG CCG AAG AAA AGC AAC GAG CTG
509 481
510 Ala Val Arg Pro Leu Arg Arg Arg Gln Pro Lys Lys Ser Asn Glu Leu
511      5              10              15
512
513 CCG CAG GCC AAC CGA CTC CCA GGG ATC TTT GAT GAC GTC CAC GGC TCC
514 529
515 Pro Gln Ala Asn Arg Leu Pro Gly Ile Phe Asp Asp Val His Gly Ser
516 20              25              30              35
517
518 CAC GGC CGG CAG GTC TGC CGT CGG CAC GAG CTC TAC GTC AGC TTC CAG
519 577
520 His Gly Arg Gln Val Cys Arg Arg His Glu Leu Tyr Val Ser Phe Gln
521      40              45              50
522
523 GAC CTT GGC TGG CTG GAC TGG GTC ATC GCC CCC CAA GGC TAC TCA GCC
524 625
525 Asp Leu Gly Trp Leu Asp Trp Val Ile Ala Pro Gln Gly Tyr Ser Ala
526      55              60              65
527
528 TAT TAC TGT GAG GGG GAG TGC TCC TTC CCG CTG GAC TCC TGC ATG AAC
529 673
530 Tyr Tyr Cys Glu Gly Glu Cys Ser Phe Pro Leu Asp Ser Cys Met Asn
531      70              75              80
532
533 GCC ACC AAC CAC GCC ATC CTG CAG TCC CTG GTG CAC CTG ATG AAG CCA
534 721
535 Ala Thr Asn His Ala Ile Leu Gln Ser Leu Val His Leu Met Lys Pro
536      85              90              95
537
538 AAC GCA GTC CCC AAG GCG TGC TGT GCA CCC ACC AAG CTG AGC GCC ACC
539 769
540 Asn Ala Val Pro Lys Ala Cys Cys Ala Pro Thr Lys Leu Ser Ala Thr
541 100              105              110              115
542
543 TCT GTG CTC TAC TAT GAC AGC AGC AAC AAC GTC ATC CTG CGC AAG CAC
544 817
545 Ser Val Leu Tyr Tyr Asp Ser Ser Asn Asn Val Ile Leu Arg Lys His
546      120              125              130
547
548 CGC AAC ATG GTG GTC AAG GCC TGC GGC TGC CAC
--> 549 TGAGTCAGCCGCCCCAGCCC 870
550 Arg Asn Met Val Val Lys Ala Cys Gly Cys His
551      135              140
552
553 TACTGCAGCC ACCCTTCTCA TCTGGATCGG GCCCTGCAGA
--> 554 GGCAGAAAACCTTAAATGC 930
555
556 TGTCACAGCT CAAGCAGGAG TGTCAGGGGC CCTCACTCTC
--> 557 GGTGCCTACTTCCTGTCAGG 990
558
559 CTTCTGGGAA TTC
560

```

base must be
in group
of 10
per 1.822 (f)
of sequence
Rules

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/319,831ADATE: 05/09/96
TIME: 16:30:54

INPUT SET: S9275.raw

Line	Error	Original Text
147	Entered (80) and Calc. Seq. Length (29) differ	(A) LENGTH: 80 base pairs
178	# of Sequences for line conflicts w/ running total	TAC GAC AGCAGCAACA ATGTAATTCT AGA
199	Entered (199) and Calc. Seq. Length (3) differ	(A) LENGTH: 199 base pairs
233	# of Sequences for line conflicts w/ running total	GCG 53
255	Entered (50) and Calc. Seq. Length (49) differ	(A) LENGTH: 50 amino acids
280	Entered (172) and Calc. Seq. Length (23) differ	(A) LENGTH: 172 base pairs
318	# of Sequences for line conflicts w/ running total	TGG 56
331	# of Sequences for line conflicts w/ running total	CAG TCC CTG GTCAGTACCT C
358	Entered (119) and Calc. Seq. Length (23) differ	(A) LENGTH: 119 base pairs
397	# of Sequences for line conflicts w/ running total	GTG 51
406	# of Sequences for line conflicts w/ running total	GTGAGTTCCG ACTCTCCTTT
428	Entered (1003) and Calc. Seq. Length (60) differ	(A) LENGTH: 1003 base pairs
549	# of Sequences for line conflicts w/ running total	TGAGTCAGCCCGCCCAGCCC 870
554	# of Sequences for line conflicts w/ running total	GGCAGAAAACCCTTAAATGC 930
557	# of Sequences for line conflicts w/ running total	GGTGCCTACTTCCTGTCAGG 990